



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

TEACHERS' MARKS AND THE RECONSTRUCTION OF THE MARKING SYSTEM¹

H. O. RUGG
University of Chicago

During the past ten years it has been increasingly evident to school men that one of the contributory causes of "failure" in the public schools has been a bad administration of the marking system. By and large in this country we fail 20 per cent of our student population. Furthermore it is relatively common, in many of our high schools at least, for teachers to fail upward of 30 per cent of their students. Recognizing that some of the difficulty may be traced to the inability of the pupil, that some of it clearly may be ascribed to a badly adjusted course of study and some of it to inefficient teaching, nevertheless there remains for the careful consideration of administrators and teachers the reconstruction and better administration of the marking system.

For just about ten years there has been much agitation in the educational press over the standardization of school marks. More than sixty articles have appeared in our various journals discussing the many phases of this problem. In 1915 the writer summarized the literature, finding at that time 39 references which dealt rather directly with the problem of the standardization of teachers' marks. Since that time no systematic summary has appeared which presented the essence of educational thought on this question or which reviewed the literature. It seems clear, therefore, that there is a need for a complete statement of present scientific thought on the matter and for the setting forth of a thoroughgoing program for the reconstruction of the marking system in our public schools. This article, therefore, is written with these three aims in view: to

¹ This article is intended to occupy a place in the list of the articles in the department of "Educational Writings" which review each month the literature in selected fields of interest.—H. O. R.

summarize the results of recent investigations, to present complete bibliographic material, and to present systematically the essence of current thinking on the problem.

I. THE MARKING SYSTEM NEEDS TO BE RECONSTRUCTED

Out of the discussion of the last fifteen years we find one point of absolute agreement, namely, that we should overhaul thoroughly the methods by which we measure the outcome of instruction in the public schools. There are three very apparent reasons for this statement: (1) the striking *variability* in teachers' marks; (2) the *unreliability*, the lack of consistency, with which teachers mark; (3) the *inconsistency in the way in which teachers distribute their marks*.

1. *The variability of teachers' marks*.—Of the 23 articles which have appeared in the past three years nearly every one presents evidence showing the striking variability in the marking of teachers. Various investigations have been reported showing the percentages of A's, B's, C's, etc., given by teachers in different systems of schools, within the same system, and within the same department. The previous article of the writer (16)¹ presented detailed evidence on this question. Such investigations as those of Starch and Elliott were quoted, in which the wide variability in judgments of teachers in marking student work was shown. Investigations reported by Chapman and Hills (5) are distinctly to the point in this connection. They show, for example, in their study of the distribution of college grades the very wide variability with which instructors even within the same department assigned the various marks of the marking scale. One instructor, for example, in several years of marking gave 400 per cent more E's than another instructor. Roberts (15) shows the same situation in the high school at Everett, Washington. Jackson (7) in a general article on the instability of teachers' marks shows the very great variability in the judgment of the examiners for the regents of the state of New York. Since we canvassed this matter so thoroughly in the earlier summary article and since the fact is so generally recognized by school people today, we need not debate it further.

¹ The figures in parentheses refer to the titles in the Bibliography at the end of this article.

2. *The unreliability of teachers' marks.*—In the same way it has become apparent to school men that teachers unaided by any objective scale will vary greatly in their judgments of the *same* student work. There are many investigations which establish this point clearly. Inglis (6), for example, in a recent article shows that of 122 high-school teachers marking an examination paper of 10 questions in plane geometry 20 per cent gave an *equal* value (10 per cent) to each question. Of the others the range in the value assigned to the various questions was from 0 per cent to 25 per cent and the average deviation was 2.4 per cent. Inglis also shows, by having the questions answered by pupils who had completed plane geometry four months before, that the relative difficulty of the 10 questions ranged from 10 to 298 points. Likewise in testing students on an algebra examination certain questions which had been marked as of equal value as others were found to be 15 and 16 times as easy as the others when worked by students. Furthermore as one studies the marks given by a teacher to the same piece of student work presented at successive times he finds even here distinct evidence of unreliability of marking. Studies referred to in the writer's previous article show clearly the following:

Under such individualistic marking systems as are in use today teachers' marks do vary widely—there are large individual differences in teachers' marks of the same students in the same subjects, on the same examination papers and the same drawings and lettering samples. The mean variations in many of the instances tested run as high as 15 per cent. They practically never are less than 5 per cent. Steele found the mean variation of eight markers on the same examination replies to vary from 2.5 per cent to 8.1 per cent. Starch found mean variation among instructors in the same department marking English papers to be 5.4 per cent. According to Starch and Elliott the probable error (almost the same as the mean variation in these cases) of 142 high-school teachers marking the same English paper was 4.5 per cent; the same geometry paper, 7.5 per cent; the same history paper, 7.7 per cent; 50 per cent of the marks were separated by more than 9 per cent. The present writer (1914) found a mean variation for teachers of lettering marking the same 20 samples of 8.6 per cent, and among students trained in lettering of 11.4 per cent, both groups never having seen an objective scale in lettering.

Starch and the present writer each found the variation in marking to be as large within any one department as that among departments and schools. Starch established a mean variation of 2.2 per cent in any one teacher's marking of the same paper. He attempts to determine the relative amounts

contributed to the total variability by the various factors affecting it. He concludes that the total difference in marking is due to differences in standards among schools, among teachers, in the same school or department, to inability to distinguish small differences in quality of school product, and to the change in an instructor's own standards. The relative amounts contributed by each may be determined from Plate 1 of the previous article (16).

The minimum size of the marking unit.—From the results of three of the four investigations bearing on this question of the size of the marking unit we may conclude that teachers, marking without an objective scale, cannot be expected to mark student work in any subject—mathematics, history, composition, lettering, etc.—within an interval of roughly 8 per cent.

3. *Inconsistency in the way in which teachers distribute their marks.*—A third type of evidence may be exhibited in showing the need for reconstructing the marking system. Not only do teachers' marks vary between themselves and within the same department, not only are they inconsistent as measures of the *same* student work, but also when they are distributed along the scale of ability we find that the shape of the distributions for different systems, for different schools, for different departments, and for different teachers is distinctly inconsistent. One teacher or school fails 30 per cent and the marks may be piled up at the low end of the marking scale. Another teacher or school gives 30 per cent of the student body marks of "excellent," or 90, etc., thereby humping up the marks at the high end of the scale. From detailed study of hundreds of mark curves it may be said that it is possible to find all forms of distribution of marks, from the form in which the marks are heavily "skewed" to the high end of the scale to the form in which the marks are piled up at the low end of the scale. There is, however, enough evidence that teachers' marks tend to be "skewed" to the high end of the scale. In the writer's previous article (16) were presented typical distribution-curves covering more than 171, 400 teachers' marks collected from 7 colleges and universities and from 39 high schools. These data certainly typify the actual condition of marking throughout the country, and conclusions made from them ought to hold for teachers at large. We may summarize the conclusions of the previous report thus:

1. Regardless of subjects of study, the typical distribution of teachers' marks results in a form or curve which may be described as skewed to the high

end of the percentile scale, i.e., the peaks of the curve will lie considerably above the mean of the base line upon which the curve is plotted and above the theoretical mean of the normal frequency-curve plotted on that base.

2. Less than 10 per cent of the several hundred mark curves, representing an aggregate of over 171,400 marks, may be described as perfectly symmetrical, and not more than two or three in a hundred of all those examined have been found to be approximately "normal."

3. The mark curves of students of advanced high-school classes are skewed much farther to the high end of the percentile scale than are the mark curves of students in corresponding elementary classes.

4.

5. There is evidence that another potent cause of the skewing is the presence of institutional "critical points" in the marking scale, e.g., the "passing" and "exemption" grades are almost always found to play an important rôle in shaping the mark curve.

II. WHY ARE TEACHERS' MARKS VARIABLE AND INCONSISTENT ?

To answer this question we proceed at once to the crux of our problem. Teachers' marks are variable and inconsistent primarily because *teachers* in marking pupils *do not measure the same trait*, secondarily because teachers *do not have a common scale* for the evaluation of definite amounts of the traits measured; i.e., their marks are distributed on a purely subjective basis and their standards are not uniform. It is possible to secure from a group of twenty-five school people from twelve to fifteen different statements as to *what* is measured by school marks. It is impossible, furthermore, to establish exactly the same connotation in the word-statements of any two teachers. Camp (3) found, for example, in canvassing this situation with his teachers that teachers said that they—

mark "improvement," "ability," "seriousness," "purpose," "moral qualities," "interest in work," "equipment." The English department, for example, marks "conscious work," "improvement and ability," while the commercial department marks "accuracy," "neatness and promptness," "honesty," "courtesy," and "seriousness." The history department marks "endeavor and improvement."

Thus we find all sorts and kinds of terms used by teachers to describe the traits that they presume to measure in assigning marks. *Ability, capacity, interest, effort, performance, achievement, accomplishment*—a chaos of marking terminology which itself shows

clearly the need for standardization, just as it also represents one aspect at least of the movement to standardize what is measured by teachers' marks. Our program then ought to include a discussion of the important question discussed in the following paragraphs.

III. WHAT IS MEASURED BY TEACHERS' MARKS?

To provide a foundation for a clear discussion of this problem we shall do well to distinguish three different traits in our student population. Let us call them for convenience *inherited capacity*, *ability-to-do*, and *specific performance*. The writer has canvassed the more than sixty articles which have appeared during the past fifteen years and found these three traits apparently fairly distinguishable among the score or more that are in the minds of teachers. With the recent development of the measuring movement, of which standardization of teachers' marks is merely one phase, there has come a general acceptance of the distinction between these three traits.

It is helpful to think of pupils as we meet them in public-school classes as being distributed over three different kinds of scale: one a scale which represents varying degrees of *inherited capacity*, another a scale which represents varying degrees of *ability*, and a third a scale which represents varying degrees of efficiency in *specific performance*.¹

Now the term *inherited capacity* practically defines itself. By it we mean the "start in life"; the sum total of nervous possibilities which the infant has at birth and to which, therefore, nothing that the individual himself can do will contribute in any way whatsoever. No one fact that has come from the study of heredity is of more importance to teachers than this fact that the capacity of an individual is determined by the "third and fourth generations" and not by anything that he himself can do. An important corollary to this statement is that an individual has a given place on the capacity scale and that only a combination of the most favorable conditions of training and work can operate to bring

¹ Limitation of space will prevent the writer from illustrating the following discussion graphically, as would be most helpful.

him to a point equally high on the ability scale. Obviously, since capacity has to do with the "start in life," in the public schools we *do not measure capacity*.

Now educational conditions (as found in the home, in the neighborhood, in the kindergarten, on the playground, in the "gang," in church, in the Sunday school, etc.) operating on "capacity" develop what we shall call "present-ability," the "*ability-to-do*," which is present at any given time. If the educational conditions are favorable, then the position of an individual on the *ability* scale will approximate his true position on the *inherited-capacity* scale. Courtis has defined capacity as "potential-ability." To express more clearly what we mean by *ability* let us discuss *specific performance*. If an individual's present-ability is his ability-to-do, then by an individual's performance we mean what he does, i.e., his *specific performance* at work. We mean, for example, the recitations that he makes, the examination papers that he writes, the machine part that he constructs, the theme that he writes, the problems that he solves, etc. *It is these specific performances of children that we measure* in school practice. It is these that school marks *ought* to measure. And it is from the sum total of these that we *estimate ability-to-do*. To reiterate an important fact, we *do not measure ability*, we *estimate it from mere performances*. An illustration will make this fact still clearer. If we give a test to measure skill in a particular school activity (such as, for example, the removal of parentheses in algebra) to a pupil five times in succession, on successive days, his specific performances may be as indicated herewith. On one day, for instance, he works in one minute 18 examples, on another day 11, on another 24, on another 19, and on a fifth 16.

There are *two important characteristics apparent in the performances of school children*. The first shows us that *specific performance is unstable* and that any *one* test does not enable us to estimate accurately the *ability* of a given pupil. The variation of this particular pupil from 11 to 24 problems a minute illustrates the variability or instability of pupils' specific performances. Furthermore it determines our attitude toward the use of performance as guides for formulating judgments on their ability-to-do.

Now the second important characteristic of a pupil's performance, and one which bears upon our question still more intimately, is that the *average of successive performances is relatively stable*. For example, in the illustrations given the average of the first two performances is 14.5, the average of the first three is 17.67, of the first four 18.0, and of the first five performances 17.60. In other words, the averages of three or more successive performances are practically stable or constant, although the individual performances vary greatly in amount. Thus it is clear that just as capacity is *potential-ability*, so *ability-to-do* may be helpfully described as the *average of a few successive performances*. Furthermore it is helpful for teachers to recall that experimentation is showing that but a very few successive performances, say three or four, need to be measured on an individual pupil to permit a close estimate of his *ability*.

We may sum up this discussion, therefore, by pointing out that if an adequate method of measuring the specific performances of pupils can be devised then the chaotic condition which is apparent in our marking terminology can easily be cleared up. Interest and effort, moral qualities, attitude, etc.—all these intangible qualities contribute to the *specific performance* of the pupil and result in making possible an estimate of his total *ability-to-do*. In this connection Preston (13) has suggested that pupil's marks ought to be stated in the form of a fraction. That writer would compare the pupil's performance with his capacity, i.e., the denominator of the fraction, for example $\frac{9}{10}$, would represent our estimate of the pupil's *ability*. The numerator of the fraction similarly would represent the teacher's judgment of the extent to which the pupil is living up to it. This Preston suggests as an administrative device which will enable us to get rather more definitely at the various phases entering into the question of marking pupils in school. Thus, in résumé, teachers' marks ought to measure ability-to-do, this in turn having been estimated from the specific performances of the pupils. We shall point out later that this estimate will be valid as we substitute objective methods for our present subjective methods of measuring ability.

IV. HOW DOES ABILITY DISTRIBUTE IN THE
GENERAL POPULATION?

Teachers' marks, if they are adequate measures of ability, ought to distribute over the ability scale in approximate conformance at least to what is known about the distribution of ability in the general population. Hence the very evident need for making use of a theoretical distribution of teachers' marks against which to check our judgment on ability as it is represented in our school classes. An extensive literature has developed concerning the form of the distribution of human traits. It has been referred to by many writers on teachers' marks and has been presented in summary fashion in a number of sources.¹ For this reason and because the writer is distinctly limited in space in the publication of this article our statement here will be very brief and seemingly dogmatic. The evidence for the statement, however, has been set forth in other writings.

The types of fact made use of by those who would standardize the marking system through the use of theoretical distribution-curves include the following: both physical and mental traits (and it is believed likewise moral traits) are found to fit reasonably well for purposes of practical application in education a mathematical curve known most commonly as the normal probability-curve. The arguments and the evidence adduced in the investigations of the past ten years lead us to believe that such a distribution-curve is reasonably adapted to either an elementary- or a secondary-school population. Granting, therefore, that we need to refer our judgment of scholastic ability to some theoretical distribution-curve and accepting this particular normal curve as a satisfactory type, we have three subordinate problems before us: (1) the determination of the number of divisions on the marking scale, (2) certain statistical points involved in the determination of the length of the scale, and (3) the determination of the number of individuals that ought to fall in the various divisions of the marking scale.

¹ For example, the reader will find a detailed statement of the statistical methods underlying the use of curves representing the distribution of human traits in the writer's *Statistical Methods Applied to Education*, chap. viii.

V. NUMBER OF DIVISIONS USED IN THE MARKING SYSTEM

In the previous article of the writer the recommendations of 9 investigators were shown with the proposed percentages for each of the divisions of the scale. At that time it was shown that all but one of these investigators recommended a five-division marking scale with the use of letters A, B, C, D, E; of words such as "excellent," "superior," "medium," "inferior," and "poor"; and in many cases of parallel percentile limits on the marking scale. Nicolson (12) has recently collected the practices of 64 colleges approved by the Carnegie Foundation. He finds that 36 of these colleges use a five-division marking scale and that among the others there is almost no central tendency. Some colleges use as few as 3 groups and others as many as 12 and 15. A large number generally accompany the use of A, B, C, D, etc., with the use of plus and minus signs. Furthermore it is shown that it is relatively common at the present time in the administration of college marking systems to translate letter marks into percentile limits on a numerical scale. For example, A=90-100, B=80-90, etc. We may sum up the situation for our colleges and high schools by pointing out that there is a fair agreement upon a five-division marking scale. We have already quoted the various suggestions made in the previous article concerning the minimum size of the unit in the marking scale. This confirms our judgment that five divisions can be handled accurately by teachers.

VI. PERCENTAGE OF INDIVIDUALS THEORETICALLY FALLING
IN EACH DIVISION

Dividing up the marking scale, therefore, into five parts we are concerned to know what proportion of human beings should be found in each of the various divisions of the scale. Without introducing a technical discussion of the statistics of the probability-curve let us say that one can obtain these facts from tables which have been constructed showing the percentages of the total area of the curve which fall between various points on the base line. A complete discussion of this matter can be found in the reference previously given. To apply the normal probability-curve rigidly to our problem would lead to a statement that if we use a five-

division marking scale the percentages of pupils that ought normally to be expected in each of the groups would be either 7, 24, 38, 24, 7, or 3.5, 24, 45, 24, 3.5, depending upon whether one breaks off the base line of the normal curve at $2\frac{1}{2}$ or 3 times the standard deviation (the latter being used merely as a unit with which to divide the base line of the curve). It is clear, however, that a sane recommendation concerning the use of a theoretical curve will lead to the statement that a percentile distribution of pupils in five divisions ought to approximate, but not necessarily to follow rigidly, any specific set of percentages. In other words, the writer would prefer to say that one would expect to find from 5 to 10 per cent A's, 20 to 25 per cent B's, 35 to 40 per cent C's, 20 to 25 per cent D's, and 5 to 10 per cent E's in each reasonably *large* group of our public-school pupils (say 100 or more). Furthermore that this device ought to be used merely as a rough check upon the total distribution of marks in various divisions of the scale for large numbers of our students and not for single classes. After assigning one's marks in a given semester the use of such a device is very helpful in making one critical of the accuracy with which he has distributed his marks.

VII. THE RECENT USE OF THE NORMAL CURVE IN DISTRIBUTING MARKS IN SCHOOLS AND COLLEGES

The literature of the past three years has revealed a number of active attempts by school and college administrators to standardize the marking system through the use of a theoretical distribution. Due to the activity of the University of Missouri, under the leadership of Professor Meyer, many colleges and schools have adopted a particular form of distribution of marks and have administered the marking system in such a way as to force a given percentage of marks of instructors into the various divisions of the scale. The use of the so-called "Missouri system," therefore, which is now in operation in such institutions as the University of Texas, reported by Benedict (2); Goucher College, reported by Kellicott (8); and Kansas Agricultural College, reported by Reisner (14), illustrates the tendency. Public-school administrators likewise can feel sure that their colleagues in the administrative world are now making

use of methods of standardization in improving the marking system. The aim in general is to cut down the variability with which teachers mark, so as to make their marks consistent in form of distribution with the abilities which they are supposed to measure. Superintendents are beginning to report systematic endeavors to get teachers to distribute their marks in rough accordance with the normal curve. A number of these attempts have made use of a five-division scheme which shows the following percentages: 5, 20, 50, 25. An article reported by Superintendent Walls, of Kent, Ohio (21), presents data illustrative of the success attained by school administrators in changing the distribution of teachers' marks from the badly "skewed" form which we commented on earlier in the article to one more closely approximating symmetry. To illustrate concretely the results of such adjustments in administration we quote from one of his tables which gives the percentile

HIGH SCHOOL

Summary of grades of six teachers. Percentage of grades assigned to each rank

YEAR	PERCENTAGE				
	Below 60 V.P.	60-74 Uns.	75-84 F.	85-94 G.	95-100 Ex.
1911-12.....	0.8	5.9	21.4	56.1	15.8
1912-13:					
First Semester.....	0.8	12.1	39.6	42.5	5.0
Second semester.....	1.1	12.0	43.6	38.0	5.3
1913-14:					
First semester.....	3.0	13.0	42.0	36.0	8.0
Second semester.....	4.0	12.0	39.0	38.0	7.0

ELEMENTARY SCHOOLS

Summary of grades of fifteen teachers. Percentage of grades assigned to each rank

YEAR	PERCENTAGE				
	Below 60 V.P.	60-74 Uns.	75-84 F.	85-94 G.	95-100 Ex.
1911-12.....	1.0	6.8	23.6	48.0	20.6
1912-13:					
First semester.....	3.4	11.7	30.0	38.4	16.5
Second semester.....	3.2	5.1	28.2	41.4	22.1
1913-14:					
First semester.....	5.0	11.0	43.0	35.0	6.0
Second semester.....	3.0	7.0	39.0	42.0	9.0

distribution of marks for five successive semesters under both his old system and his new one. The writer has had submitted to him in manuscript form other distributions showing the same sort of procedure with, in general, the same sort of results.

VIII. THE NEED FOR CLEAR WORD-DESCRIPTIONS OF "ABILITY" IN OUR MARKING SYSTEM

These proposed reforms have to do with the general adequacy, with the general validity, of the teacher's measurement of pupil-ability. Thus far the workers in this field have proposed the standardization of the marking system primarily through the use of theoretical distribution-curves which are believed to represent the distribution of ability in the general population. What we have done in the movement up to date is to translate our desire for scale measurement in marking pupils into a system of numbers, or into a system of letters, without defining clearly what the letter or number represents. Now in the practical adaptation of a marking scale either letter or numerical, an important consideration faces the administrator: how shall various degrees of *ability-to-do* in school children be defined?

It seems quite clear that neither the number nor the letter has carried a common meaning to the student, to the teacher, and to the administrator alike. We must not forget that these three groups, students, teachers, and administrators, have to "think together" concerning school marks, and our aim is to build a marking system simple and economical to administer, and yet one which will enable these three sets of minds to agree on the marks to be put on the results of instruction. Now the common avenue of communication is words. We think together only when we use words in communicating, about whose connotation there is no doubt. Hence in our practice of using letters and numbers we have forsaken a worded system of measures or marks and have made use of a set of letters or, worse still, a set of numbers whose meaning has not been interpreted in commonly used words. Hence the differences in interpretation; hence the bickering about 79, 78.3, 81, 80, etc.; hence the ridiculous, yes criminal, practice of failing children at 69 or 68 when the passing mark is 70!

If this is one of the chief causes of our marking difficulties the ultimate remedy stands out clearly: *substitute clear word-statements for abstruse literal and numerical symbolism*. First, the marking system for each subject of study should consist of a set of word-descriptions of "ability" for various grades of students; secondly, these statements, these standards, should be expressed in words common to the thinking of pupil, teacher, and administrator; thirdly, they should be brought to the attention of each group constantly enough to result in thorough acquaintance with the marking standards and in relatively accurate use of them. That is, they should be printed, charted, and placed before teacher and pupil constantly. If not in the textbook, then in charts on the walls of the classroom, in printed statements to be pasted into the textbook, in printed folders to be sent home to parents. Instead of being told that John's mark is 79 in mathematics or 92 in English, pupils and parents alike should be told that John's mark represents the ability to do such and such things, to write equal to 60 on the Ayres scale, to spell successfully a given proportion of the eighth-grade words in such and such a list, etc. Definiteness of understanding on the part of all concerned should be the outcome of the setting up of such types of standards. It is a difficult thing to do, this writing out of word-statements of degrees of ability. It is, however, one of the most important co-operative tasks of teacher and administration.

Thus it will be possible to define relatively accurately the performances (and from these to estimate the ability) of a student in the top fifth, or of a "failure," a "good," a "superior," or a "mediocre" student. For convenience, of course, we may find it desirable to use letters or even numbers to typify, to represent economically the degrees of ability—but the foundation of the scheme is the "*word-description*." Furthermore it is difficult to see how any sound system can be installed without using absolutely objective measures as illustrated above. A "passed" student in handwriting in the elementary school is one whose ability-to-do is represented by a particular sample on one of the standard handwriting scales (implying training on the part of both teacher and

pupil in judging handwriting on objective scales, and hence fair agreement in measurement and in interpretation of the same piece of student work).

IX. THE IMPORTANCE OF "RANKING" PUPILS IN ORDER OF ABILITY

There are two possible approaches to the problem of reconstructing our marking practice. One involves the use of an absolute definition of ability, the other the use of measurement by relative position.¹ The foregoing proposal to define grades of ability in carefully worded statements and separately for each subject of study represents, in the writer's opinion, the best possible practice that schools can adopt today. Involved in this application of standardizing methods in the improvement of the marking system there is, however, an important subordinate step. Teachers and administrators can and ought to make use of the helpful practice of *marking pupils by ranking them first in order of ability*, following up the ranks thus obtained by approximate assignment to the various letter or number groups of the marking scale. This in turn should be checked up by distributing pupils to different groups by the use of some such percentile allotment of the normal distribution as 5 to 10, 20 to 25, 35 to 40, 20 to 25, and 5 to 10 for successive

¹ *A proposed method of using the 100 per cent scale.*—The writer wishes to suggest in this connection that if a school system cannot bring itself to give up a detailed number scale which distinguishes pupils in terms of specific numerical values on a 100 per cent scale, common agreement in the thinking of pupil and teacher would be enhanced by using the *whole* percentile scale from 0 to 100. Present practice in American schools makes use of about 0.3 to 0.4 of the 100 per cent scale, the passing mark being placed *arbitrarily* at 70, 75, or 60. Teachers are thereby forced to distribute the members of their classes by separate numerical measures over a scale of 25 or 30 units.

To use consistently the whole 100 per cent scale would imply either the *arbitrary* placing of division points on the scale (at the present time it is done in this fashion, the various marking divisions often being unequal in length, for example, 70 to 80, 80 to 90, 90 to 95, 95 to 100) or the determination of these points by reference to the distribution of individuals over the base line of a distribution-curve. The latter method can now be used, taking the base line of the normal probability-curve as a scale of ability and dividing it into five divisions (or the number that our thinking leads us to take as most practicable).

The placing of the passing mark.—If we assume the normal curve as the foundation of our distribution of ability and then divide the marking scale into a given number

fifths of the scale. Lacking an approximate statement of absolute abilities such as would be represented by a detailed word-statement of various grades of ability the teacher can rank pupils in order of ability, at least in descending *groups* of ability, and thereby measure them by their relative position in the total groups. This relative position can, closely enough for practical purposes and also fairly easily, be turned into definite marks by assigning certain approximate percentages of the entire pupil population to various groups, deciding the "border-line" cases perhaps on subjective judgment. At least the rough use of the distribution-curve will in this case result in a more critical attitude on the part of the teacher toward her own judgment and in the long run in a sounder measurement of ability in her classes.

X. A PROGRAM FOR THE RECONSTRUCTION OF THE MARKING SYSTEM

As a summary statement of the present status of thought the following steps are both desirable and practicable as means of thoroughly rebuilding our public-school marking system:

1. The evident lack of reliability and consistency in teachers' marks and their evident inaccuracy as measures of ability demand of administrators the *initiation of campaigns of education among*

of parts we automatically establish either one of two important facts, (1) the value of the passing mark and (2) the proportion of pupils who ought to fall in the various divisions of the scale. If we make use of the whole 100 per cent scale, assume the normal curve as representing the distribution of ability, assume the proportion of our group of pupils above which we shall not permit ourselves to fail students, then we find our passing mark locating itself somewhere in the region of 15 to 20 on the scale. That is to say, instead of setting up an arbitrary number, such as 70 or 75, which has no direct meaning in terms of pupil-ability, we now set up an upper numerical limit concerning the number of pupils who evidently are unable to master the essentials of our courses of study. If we take as this upper limit, say, 5 to 10 per cent, perhaps 7 or 8, then examination of the area of the probability-curve shows us that if we go over, say, 20 points on the base line from the 0 point at the left we have included roughly 10 per cent of the area of the curve. If our "limit" of failure is smaller, then our "passing mark" will be lower. Furthermore the specific location of the position of any pupil on this scale of ability may be defined rigidly in terms both of relative position and of absolute numerical score. An individual who receives a mark of 75, for example, is now an individual located definitely with respect to rank in the group, provided the group distributes in ability in accordance with the area of the curve which has been used.

their teachers to a recognition of the importance of the facts set forth. In the carrying on of such campaigns helpful administrative devices have been found to be:

a) The publication of the distribution of teachers' marks each semester in open bulletins.

b) The discussion of these bulletins in teachers' meetings.

c) Insistence that each teacher tabulate and plot graphs of her distribution of marks each semester before submitting them to the office.

d) Requiring reading and discussion of the use of distribution-curves in marking.

e) Insistence that each teacher rank her pupils prior to assigning final marks, whether on examination, "paper," "quiz," or semester's work.

f) The appointment of departmental committees with instruction to define, *in detailed word-statements*, each grade of ability represented on the marking scale.

g) The use of objective scales and tests in all those subjects and for all those types of subject-matter for which such tests and scales are now available.

h) The use of "general-ability" tests for purposes of classifying pupils and of detecting various grades of ability in our pupils early in the course of instruction.

2. If letters are now used in a school system certainly each one should be merely a symbol to signify the abilities-to-do which have been defined in very detailed worded statements, understood alike by teacher, administrator, and pupil, in terms of which the instruction has constantly been oriented.

3. If numbers are used they should be employed only as economical symbols to represent the various groups or divisions of the marking scale; that is, only one number (for example, the median one) should be used to typify a group of marks (as, for example, the "excellent" group, the "A" group), instead of using, as is now common, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100. Thus the only numbers used would be such clearly separated numbers as 95, 90, 85, etc. Numbers in themselves would cease to have the "picayunish" absolute differences in meaning that they pretend to now and

would merely stand for abilities which have been distinguished clearly through detailed worded statements.

4. As a practical and helpful tool in the adequate measurement of student work, measurement by ranking with subsequent transmutation to absolute marks by means of a distribution-curve (preferably the normal probability-curve) is very desirable.

5. Probably few of our school administrators will feel that they can afford to change their 70 per cent passing-mark to one of 20 per cent or 15 or 25 per cent. In connection with the suggestion made in the footnote on p. 715 the writer would suggest that if a general overhauling of the whole numerical scheme could be made it would doubtless be a helpful revision from the standpoint of practical interpretation.

XI. BIBLIOGRAPHY

The following articles and monographs were published on the subject of teachers' marks during the period January, 1915, to January, 1918. Reference No. 17 in the following list gives a complete bibliography (39 items) to January, 1915.

1. Albright, G. H. "How Teachers Mark," *School and Society*, III (1916), 462.
2. Benedict, H. Y. "Standardization of Grades at the University of Texas," *School and Society*, III (1916), 105.
3. Camp, Frederick. "Some 'Marks': an Administrative Problem," *School Review*, XXV (December, 1917), 697.
4. Canning, J. B. "The Meaning of Student Marks," *School Review*, XXIV (1916), 196-202.
5. Chapman, J. Crosby, and Hills, Myra E. "A Statistical Study of the Distribution of College Grades," *Pedagogical Seminary*, XXIII (1916), 204-10.
6. Inglis, Alexander. "Variability of Judgments in Equating Values in Grading," *Educational Administration and Supervision*, II (1916), 25-31.
7. Jackson, Nelson A. "Grading Papers in Geometry," *School Science and Mathematics*, XVII (1917), 483.
8. Kellicott, W. E. "Grading at Goucher," *Science*, N.S., XLI (1915), 909.
9. Kelly, F. J. "Teachers' Marks," Teachers College, Columbia University, *Contributions to Education*.
10. Maxwell, G. E. "The Grading of Students," *School and Society*, VI (June 28, 1917), 113.
11. Miller, Wade E. "The Interpretation of School Grades," *Pedagogical Seminary*, XXIV (1917), 384-90.

12. Nicolson, Frank W. "Standardizing the Marking System," *Educational Review*, LIV (1917), 225-37.
13. Preston, Carolyn E. "Are Our Schools Hitting the Mark?" *Educational Review*, LI, 275-85, esp. 281 ff.
14. Reisner, Edward H. "Grading System Adopted by the Kansas State Agricultural College to Go into Effect September, 1916," *Educational Administration and Supervision*, III (September, 1917), 419.
15. Roberts, Alexander C. "A Study of the Marking System of the Teachers of the Everett High School," *Educational Administration and Supervision*, III (October, 1917), 485.
16. Rugg, H. O. "Teachers' Marks and Marking Systems," *Educational Administration and Supervision*, I (1915), 117-42.
17. Sechrist, Frank K. "The Process of Examining," *Educational Review*, L (1915), 399-417.
18. Smith, R. C. "Teachers' Marks in High School," *American School Board Journal* (November, 1917), p. 18.
19. Starch, D. "Can the Variability of Marks Be Reduced?" *School and Society*, II (1915), 242.
20. Van Cleave, J. H. "Distribution of Grades in an Elementary Zoölogy Course in the University of Illinois," *Educational Administration and Supervision*, II (1916), 251-53.
21. Walls, W. A. "Variability of Grades Assigned to Pupils by Teachers," *American School Board Journal*, L-LI (January, 1915), 8.
22. Weld, L. D. "A Standard of Interpretation of Numerical Grades," *School Review*, XXV (June, 1917), 412.
23. Witham, Ernest C. "Teachers' Marks in Eighth A, B, and C Division," *Educational Administration and Supervision*, II (1916), 243-51.